様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ５ 月 ２６ 日現在

機関番号：１２６０１
研究種目：若手研究
研究期間：2019～2022
課題番号：１９Ｋ２０３７０

研究課題名（和文）実ロボットにおける自律的な軌道計画を実現する階層型深層強化学習の開発

研究課題名（英文）Hierarchical Reinforcement Learning for Autonomous Motion Planning with Real Robots

研究代表者
長　隆之（Osa, Takayuki）

東京大学・大学院情報理工学系研究科・准教授

研究者番号：５０８０４６６３

交付決定額（研究期間全体）：（直接経費）　　4,600,000円

研究成果の概要（和文）：強化学習は，実社会で自律的に動くロボットを実現するためのアプローチとして期待される一方，学習効率や環境の変化への適応などに課題を抱えている。本研究は，様々な動きを学習し，それらを使い分けることにより，環境の変化へと適用できる深層強化学習の枠組みを構築することを目指した．まず，ロボットの動作計画に必要な軌道最適化の問題において，多様な解を発見するアルゴリズムを構築した。また，その知見を活かし，深層強化学習においても，多様な挙動を発見・モデル化するアルゴリズムを構築した．また，提案するアルゴリズムによって得られた多様な挙動を使い分けることによって，環境の変化への適応を効率的に行えることを示した。

研究成果の学術的意義や社会的意義
従来の研究において，ロボットの動作計画問題には無数の多様な解が存在しうることが指摘されていたが，それらを一括して導出しモデル化する手法はこれまでなかった。本研究の成果は，無数の多様な軌道を一括して導出・モデル化することを可能にした点で新規性が高い。同様に，多様な挙動を一度に学習することを深層強化学習においても実現した点にも価値がある。本研究で得られた成果は，強化学習等を活用したロボットシステムにおいて環境の変化への適応を劇的に効率化する可能性を秘めており，実社会での適用先を広げると考えられる。また，これらの成果は国際的に認知され，2022年にはロボット学習分野のトップ学会にて招待講演を行った。

研究成果の概要（英文）：Reinforcement learning (RL) is a promising approach to realizing autonomous robots that work in the real world. However, RL faces challenges in learning efficiency and adaptation to changes in the environment. This study aimed to develop a framework for deep RL that can be adapted to changes in the environment by learning various types of movements and using them differently. First, we constructed an algorithm that finds diverse solutions to the problem of trajectory optimization, which is necessary for robot motion planning. Using this knowledge, we also developed an algorithm for deep RL that can find and model a myriad of solutions. We also showed that the algorithm can efficiently adapt to changes in the environment by using a variety of behaviors obtained by the proposed algorithm.

研究分野：知能ロボティクス

キーワード：強化学習　軌道計画　ロボティクス

１．研究開始当初の背景

強化学習は，試行錯誤を通じて最適な戦略を学習するという機械学習の一分野である．深層学習を用いたものとは深層強化学習と呼ばれ，囲碁などのゲームにおいて人間を上回る性能を示したことで，社会的な注目を集めるようになった．この強化学習はロボットへの適用も進められており，今後大きな性能の飛躍が期待されている．しかしその一方で，深層強化学習には様々な課題が存在し，それらが要因となって実世界への適用範囲が狭いものとなっている．実世界で強化学習が機能するためには，より少ない手間で自律的・効率的に動作を獲得する，異なる状況に素早く適応する，といった機能が求められる．強化学習の学習効率を上げるためのアルゴリズムの工夫についてはこれまで多くの研究が行われているものの，実世界での強化学習の適用範囲を劇的に広げるような成果は見られていない．これらの研究においては，ランダムに初期化されたニューラルネットワークをいかに効率よく訓練するか，ということに主眼があった．

一方で，人間がタスクを学習する過程に目を向けてみると，人間自身は，必ずしも全く経験や知識がない状態から動きを学習しているわけではない．人間は，過去に獲得した動作のレパートリーを用い，それらを変化させる，あるいは組み合わせることで，新たな動きを素早く獲得していると考えられる．このような枠組みを強化学習にも取り入れることにより，強化学習の実用性を高めることが期待できる．

２．研究の目的

そこで本研究は，様々な動きを学習し，それらを使い分けることにより，環境の変化へと適用できる深層強化学習の枠組みを構築することを目指した．具体的には，1)一つのタスクを実行するための多様な解を発見，モデル化するアルゴリズムを構築し，2)それらをオプションとして使い分けることによって，環境の変化への適用などを実現する枠組みの構築に取り組んだ．

既存の研究において，動作のレパートリーを使い分けることによって，与えられたタスクを効率よく解くという枠組みは長年研究されてきている．しかし，動作のレパートリー自体は，エンジニアによって事前にデザインされていたり，エキスパートによるデモンストレーションを事前に計測しておいて学習に利用するなど，開発者が時間と手間をかけて有用な挙動を準備しているものが多かった．しかし，人間をはじめとする生物の学習過程においては，自律的な学習や探索をへて，有用な動作のレパートリーが獲得される．ロボットにおいても，同様に自律的に多様な動作を獲得できることが望ましい．本研究では，自律的に多様な挙動を獲得するアルゴリズムを構築するという点に注力して研究を行うこととした．

３．研究の方法

本研究では，多様な解を一つのモデルで表現するにあたり，潜在変数を入力とするモデルを用い，潜在空間および潜在変数を入力とするモデルをどのように学習するかということに焦点を置いて研究を行った．まずは軌道最適化の問題に取り組み，複数の解を見つけ出すということをどのように定式化できるか，そしてどのようなアルゴリズムに落とし込むことができるかについて検討した．具体的には，複数の解を見つけ出す問題を，多峰性を持つ分布の密度を推定する問題に落とし込み，元の最適化問題と，密度推定問題の数理的な関係について検討した（図 1）．

その次の段階として，多様な解を発見・モデル化する深層強化学習アルゴリズムの開発に取り組んだ．多様な解を発見する深層強化学習を解の潜在空間を学習する問題として捉え，相互情報量の最大化を通して多様な解を学習するアルゴリズムを開発した．この際，相互情報量に関係する目的関数を直接方策をモデル化するニューラルネットワークに誤差伝播できるような枠組みにすることで，安定した学習性能が得られる手法を検討した．これらの成果を用い，多様な解を発見することで，新たな環境へと適応できる枠組みを構築した．

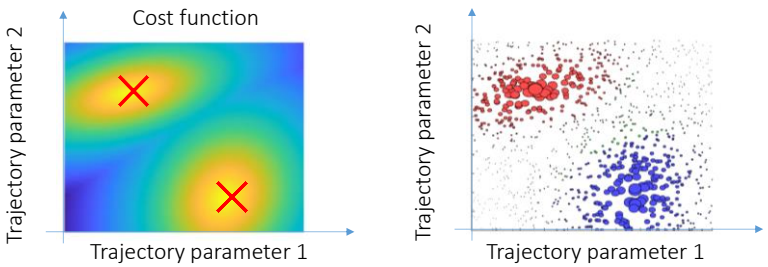

図 1. 左図のコスト関数を最小化する問題を右図の密度を推定する問題に置き換えることで複数の解を見つけ出すアルゴリズムを構築した [Osa IJRR 2020].

４．研究成果

(1)軌道最適化における多様な解を発見するアルゴリズム

軌道最適化の問題において，2020 年度に複数の解を導出する軌道最適化アルゴリズムを構築した．この研究において，コスト関数の最小化とボルツマン分布に基づく密度推定の関係性を数理的に示し，複数の解を得るための実用的なアルゴリズムを提案している．当初開発したアルゴリズムは有限個の解を発見するものであったが，2022 年度には無数の解をモデル化し，連続的に軌道の形状を変化させることのできる軌道最適化アルゴリズムを構築した（図 2）．また，環境に追加で障害物などが設置された際にも，学習された多様な軌道の中から適切な軌道を選び出すことで，高速に軌道計画ができることを実験的に示した．このアプローチは，学習された挙動の低次元な潜在空間の中で解を探索することで，効率よく軌道を計画していると解釈できる．

これらの成果は，ロボティクス分野におけるトップレベルの学術誌として知られる International Journal of Robotics Research にて発表された[Osa, 2020] [Osa, IJRR, 2022].

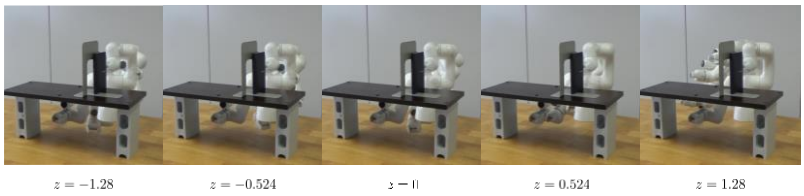

図 2 軌道最適化において，多様な解を学習した例[Osa IJRR 2022].

（２）複数の解を発見する深層強化学習アルゴリズム

軌道最適化において得られた知見を活かし，無数の解を発見・モデル化する深層強化学習アルゴリズムを構築した．構築した深層強化学習では，連続的な潜在空間を学習することで，多様な歩行挙動を一つのモデルで表現した上で，連続的に挙動のタイプを変化させることができる(図 3)．さらに，このように学習した多様な歩行挙動を用いることで，歩行するエージェントの身体に一部変更を加えた場合でも，少ない試行で適応できることを実験的に示した．このアプローチは，軌道計画における環境の適応と同様に，低次元な潜在空間の中で適切な挙動を探索していると解釈することができる．この成果は[Osa et al. Neural Networks 2022]として発表した．これらの研究成果は国際的にも認知され，2022 年 12 月の Conference on Robot Learning にて招待講演を行った．

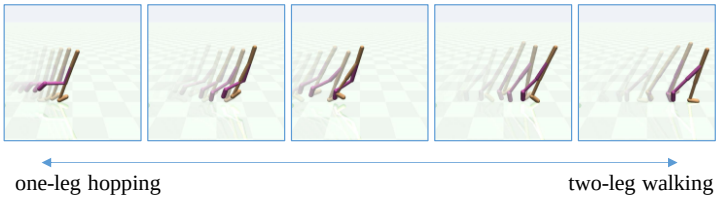

図 3 多様な歩行挙動を連続的な潜在空間を用いて学習した例 [Osa et al. Neural networks 2022].

（３）複数の挙動を用いることによるマルチエージェント強化学習における方策のロバスト化

自律的に学習される多様な挙動を活用する一つの方向性として，協調型マルチエージェント強化学習における方策のロバスト化を行う枠組みを構築した．協調型マルチエージェント強化学習においては，複数のエージェントが存在し，試行錯誤を通してお互いが協調して動作する方法を学習するが，協調動作する相手方の方策に変化があると，とたんに方策が機能しなくなる場合があることが知られている．そこで，本研究で構築した多様な方策を自律的に学習するアルゴリズムを拡張し，相手エージェントの多様な方策に対してロバストに機能する方策を得る枠組みを構築した．また，提案手法においては，方策の学習に敵対的学習の考え方を組み込むことで，多様な方策を用いるだけでは得られないロバスト性を実現している．このアルゴリズムはロボットによる介護タスクに適用され，実用性が示された．この成果は現在国際会議論文として投稿し，査読中である．

<発表文献>

① Takayuki Osa, Multimodal Trajectory Optimization for Motion Planning, The International Journal of Robotics Research, Vol. 39 No. 8, pp. 983―1001, 2020.
② Takayuki Osa, Motion Planning by Learning the Solution Manifold in Trajectory Optimization, The International Journal of Robotics Research, Vol. 41, No. 3, pp. 291―311, 2022.
③ Takayuki Osa, Voot Tangkaratt, Masashi Sugiyama, Discovering Diverse Solutions in Deep Reinforcement Learning by Maximizing State-Action-Based Mutual Information, Neural Networks, Vol. 152, pp. 90―104, 2022.

５．主な発表論文等

〔雑誌論文〕　計7件（うち査読付論文　6件／うち国際共著　2件／うちオープンアクセス　1件）

１．著者名 Osa Takayuki、Tangkaratt Voot、Sugiyama Masashi	４．巻 152
２．論文標題 Discovering diverse solutions in deep reinforcement learning by maximizing state?action-based mutual information	５．発行年 2022年
３．雑誌名 Neural Networks	６．最初と最後の頁 90～104
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.neunet.2022.04.009	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Osa Takayuki	４．巻 41
２．論文標題 Motion planning by learning the solution manifold in trajectory optimization	５．発行年 2022年
３．雑誌名 The International Journal of Robotics Research	６．最初と最後の頁 281～311
掲載論文のDOI（デジタルオブジェクト識別子） 10.1177/02783649211044405	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Osa Takayuki、Aizawa Masanori	４．巻 10
２．論文標題 Deep Reinforcement Learning With Adversarial Training for Automated Excavation Using Depth Images	５．発行年 2022年
３．雑誌名 IEEE Access	６．最初と最後の頁 4523～4535
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/ACCESS.2022.3140781	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 長 隆之	４．巻 39
２．論文標題 ロボットのための階層型深層強化学習	５．発行年 2021年
３．雑誌名 日本ロボット学会誌	６．最初と最後の頁 613～616
掲載論文のDOI（デジタルオブジェクト識別子） 10.7210/jrsj.39.613	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

1．著者名 Takayuki Osa, Shuhei Ikemoto	4．巻 1
2．論文標題 Goal-Conditioned Variational Autoencoder Trajectory Primitives with Continuous and Discrete Latent Codes	5．発行年 2020年
3．雑誌名 SN Computer Science	6．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s42979-020-00324-7	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

1．著者名 Takayuki Osa	4．巻 39
2．論文標題 Multimodal Trajectory Optimization for Motion Planning	5．発行年 2020年
3．雑誌名 The International Journal of Robotics Research	6．最初と最後の頁 1,19
掲載論文のDOI（デジタルオブジェクト識別子） 10.1177/0278364920918296	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1．著者名 Hiroyuki Karasawa, Tomohiro Kanemaki, Kei Oomae, Rui Fukui, Masayuki Nakao, Takayuki Osa	4．巻 5
2．論文標題 Hierarchical Stochastic Optimization with Application to Parameter Tuning for Electronically Controlled Transmissions	5．発行年 2020年
3．雑誌名 IEEE Robotics and Automation Letters	6．最初と最後の頁 628, 635
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/LRA.2020.2965085	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

〔学会発表〕　計7件（うち招待講演　3件／うち国際学会　4件）

1．発表者名 Takayuki Osa
2．発表標題 Discovering diverse solutions in reinforcement learning
3．学会等名 Workshop on Functional Inference and Machine Intelligence（招待講演）（国際学会）
4．発表年 2023年

１．発表者名
Takayuki Osa

２．発表標題
Dealing with the objective function with multiple extrema in robot learning

３．学会等名
Conference on Robot Learning（招待講演）（国際学会）

４．発表年
2022年

１．発表者名
長隆之

２．発表標題
多様なロボットの挙動を学習する深層強化学習

３．学会等名
第39回日本ロボット学会学術講演会

４．発表年
2021年

１．発表者名
長隆之

２．発表標題
解の潜在空間を用いた軌道計画

３．学会等名
第38回日本ロボット学会学術講演会

４．発表年
2020年

１．発表者名
Johannes Ackerman, Takayuki Osa, Masashi Sugiyama

２．発表標題
Reducing Overestimation Bias in Multi-Agent Domains Using Double Centralized Critics

３．学会等名
NeurIPS 2019 Deep Reinforcement Learning Workshop（国際学会）

４．発表年
2019年

１．発表者名 Takayuki Osa
２．発表標題 How should we design a robot learning system?
３．学会等名 Workshop on Robot Learning: Control and Interaction in the Real World, NeurIPS 2019（招待講演）（国際学会）
４．発表年 2019年

１．発表者名 Takayuki Osa
２．発表標題 Trajectory optimization via density estimation
３．学会等名 第37回日本ロボット学会学術講演会
４．発表年 2019年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関